JOURNAL OF VIROLOGY, Nov. 1994, p. 7467–7481
0022-538X/94/$04.00+0

Vol. 68, No. 11

Genetic Differences between Blood- and Brain-Derived Viral Sequences from Human Immunodeficiency Virus Type 1-Infected Patients: Evidence of Conserved Elements in the V3 Region of the Envelope Protein of Brain-Derived Sequences

BETTE T. M. KORBER,[1,2] KEVIN J. KUNSTMAN,[3] BRUCE K. PATTERSON,[3] MANOHAR FURTADO,[3] MIRANDA M. McEVILLY,[1] ROBERT LEVY,[4] AND STEVEN M. WOLINSKY[3]*

Theoretical Biology and Biophysics (T10), Theory Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545[1]; Santa Fe Institute, Santa Fe, New Mexico 87501[2]; and Department of Medicine[3] and Department of Neurosurgery,[4] Northwestern University University Medical School, Chicago, Illinois 60611

Received 4 March 1994/Accepted 26 July 1994

Human immunodeficiency virus type 1 (HIV-1) sequences were generated from blood and from brain tissue obtained by stereotactic biopsy from six patients undergoing a diagnostic neurosurgical procedure. Proviral DNA was directly amplified by nested PCR, and 8 to 36 clones from each sample were sequenced. Phylogenetic analysis of intrapatient envelope V3-V5 region HIV-1 DNA sequence sets revealed that brain viral sequences were clustered relative to the blood viral sequences, suggestive of tissue-specific compartmentalization of the virus in four of the six cases. In the other two cases, the blood and brain virus sequences were intermingled in the phylogenetic analyses, suggesting trafficking of virus between the two tissues. Slide-based PCR-driven in situ hybridization of two of the patients' brain biopsy samples confirmed our interpretation of the intrapatient phylogenetic analyses. Interpatient V3 region brain-derived sequence distances were significantly less than blood-derived sequence distances. Relative to the tip of the loop, the set of brain-derived viral sequences had a tendency towards negative or neutral charge compared with the set of blood-derived viral sequences. Entropy calculations were used as a measure of the variability at each position in alignments of blood and brain viral sequences. A relatively conserved set of positions were found, with a significantly lower entropy in the brain- than in the blood-derived viral sequences. These sites constitute a brain "signature pattern," or a noncontiguous set of amino acids in the V3 region conserved in viral sequences derived from brain tissue. This brain-derived signature pattern was also well preserved among isolates previously characterized in vitro as macrophage tropic. Macrophage-monocyte tropism may be the biological constraint that results in the conservation of the viral brain signature pattern.

One of the devastating complications of human immunodeficiency virus type 1 (HIV-1) infection is the associated impairment of neurological function in many individuals. Signs and symptoms of HIV-1 encephalopathy include marked cognitive impairment in many spheres of intellectual functioning, motor abnormalities, and, in children, developmental delay or regression in intellectual and motor milestones (6, 7, 62). In both children and adults, such encephalopathy can be due to direct infection of the brain by HIV-1, as evidenced by the absence of secondary opportunistic infections or neoplasms. Neuropathological lesions in HIV-1-infected brains include multiple disseminated foci of microglia, macrophages, and multinucleated giant cells (encephalitis); diffuse white matter damage with myelin loss, reactive astrogliosis, macrophages, and multinucleated giant cells (leukoencephalopathy); and diffuse reactive astrogliosis and microglial activation with neuronal loss (diffuse poliodystrophy) (7, 62). Ultrastructural and in situ hybridization studies have shown that the predominant cells consistently infected by HIV-1 within the brain are microglia and macrophages (17, 36, 78), which express the CD4 surface molecule required for virus binding to the cell surface (34, 75). HIV-1-specific proteins and nucleic acid sequences have rarely been detected unambiguously in endothelial cells, astrocytes, oligodendrocytes, or neurons (49, 78). Primary infection of microglia and macrophages is further supported by macrophage tropism of viral isolates derived from the central nervous system (CNS) (9, 10, 71, 74).

Data from a number of laboratories suggest that some of the molecular determinants of cellular tropism as well as replication efficiency and cytopathicity are localized within the V3 loop of gp120 (11, 13, 23, 33, 64, 76, 77). The V3 loop also plays a role in virus-cell absorption and fusion following gp120 binding to the CD4 molecule on the target cell surface. Conservation of amino acids at selected positions in the V3 loop has been found by comparison among epidemiologically unlinked laboratory-adapted macrophage-tropic strains (2, 3, 24, 25). Amino acid substitutions in the V1 and V2 regions of gp120 (1, 5, 28) and in gp41 (40) also influence the biological properties of the virus. In general, these data, combined with studies of ligand binding affinity and neutralization efficiency (52, 55, 60, 61, 81), suggest that specific combinations of substitutions interspersed throughout the envelope coding region may impose a conformational change in the gp120 molecule that can alter the biological phenotype of the virus.

* Corresponding author. Mailing address: Dept. of Medicine, Northwestern University Medical School, Olson 8427, 710 N. Fairbanks, Chicago, IL 60611. Phone: (312) 908-9819. Fax: (312) 908-4588. Electronic mail address: steve@elvira.cac.nwu.edu.

TABLE 1. CD4 counts and diagnoses

Patient no.	CD4 count (cells/mm^3)	Diagnosis
1	39	Lymphoma
2	8	Lymphoma
3	110	Toxoplasmosis
4	124	Progressive multifocal leukoencephalopathy
5	249	Necrotizing encephalitis
6	30	Progressive multifocal leukoencephalopathy (first biopsy); toxoplasmosis and *Mycobacterium avium-intracellulare* (second biopsy)

To analyze genetic patterns in brain- and blood-derived viruses, proviral DNA sequences from six patients were directly amplified from viable, uncultured brain biopsy tissue and compared with proviral DNA sequences from simultaneously obtained blood samples. Analysis of these sequences demonstrates that viral sequence heterogeneity exists in different body compartments within an infected individual and supports evidence that charge and conformation changes in the V3 loop may result in differences in cell tropism (23). A characteristic "brain signature pattern" was identified in the V3 region based on particular amino acid positions that were highly conserved among brain-derived sequences from the six patients but variable in the blood-derived sequences from the same six patients. This pattern was objectively defined by entropy measurements of aligned amino acid positions. This brain signature pattern, based on samples directly amplified from viable tissue, confirmed and extended the number of previously observed amino acid restrictions associated with macrophage tropism in vitro (11, 13, 23, 33, 64, 76, 77).

MATERIALS AND METHODS

Study subjects and clinical procedures. Clinical material was obtained from six HIV-1-infected patients with significant neurological signs and symptoms requiring image-guided stereotactic brain biopsy for definitive diagnosis. A clinical history, neuropsychiatric tests, and CD4$^+$ lymphocyte counts were obtained for each study subject. Neurological signs and symptoms were consistent with the onset of global neurological dysfunction, with clinical evidence supporting acute rather than chronic HIV-1-associated neurological disease. The patients most commonly presented with generalized seizures (four of six) and changes in mentation (two of six). No patient had neurological problems prior to the presenting event. Each patient's signs and symptoms resolved after specific therapeutic intervention was initiated.

All patients were from the United States and had low CD4$^+$ lymphocyte counts (Table 1). The samples were taken in 1990 from patients 1 to 4 and in 1992 from patients 5 and 6. Peripheral blood was collected in preservative-free heparin during the operation. All samples were coded to preserve the anyonymity of the study subjects.

Contiguous brain biopsy cores (2 by 20 mm) were immediately placed in sterile tubes containing 4 M guanidium isothiocynate for nucleic acid analysis, Dulbecco's modified Eagle's medium (DMEM) for virus isolation, formalin for histopathological examination, and STE (Streck Labs, Omaha, Neb.) for in situ hybridization analysis. The brain tissue samples in 4 M guanidium isothiocynate and DMEM were frozen and stored at −135°C. Brain biopsy specimens were also processed for routine histopathological and virological evaluation as described previously (48). Diagnostic studies were performed for microbiology (viruses, including cytomegalovirus [CMV], herpes simplex virus, and polyomavirus; bacteria; and protozoa), cytology, and histopathology as described previously (48).

Nucleic acid preparation. Peripheral blood mononuclear cells (PBMCs) were obtained from whole blood by discontinuous Ficoll-Hypaque density gradient centrifugation. The PBMCs were washed in phosphate-buffered saline, and the cell concentration was measured with a hemacytometer. Approximately 10^6 cells were resuspended in 400 μl of 10 mM KCl–10 mM Tris-HCl (pH 8.3)–2.5 mM $MgCl_2$–0.5% Tween 20–0.5% Nonidet P-40–100 μg of proteinase K per ml and incubated at 55°C for 1 h. The cell DNA was incubated at 95°C for 10 min to heat-inactivate the proteinase K.

The brain biopsy cores were dispersed in 4 M guanidium isothiocynate with an Omnimix homogenizer in a disposable vessel. Total nucleic acids were extracted with phenol-chloroform (24:1, chloroform-isoamyl alcohol), ethanol precipitated, and resuspended in deionized water. This method minimizes the systematic bias towards one genetic variant, as is observed when the virus is cultured (45, 72).

Oligonucleotide primers and probes. The HIV-1 oligonucleotide primers and probes were designed using the HIV-1 consensus sequences contained in the Los Alamos Human Retroviruses and AIDS database (56). Oligonucleotides were synthesized on an Applied Biosystems 380B DNA synthesizer using phosphoramidite chemistry. The synthesized material was alkaline deprotected and purified by elution through an oligonucleotide purification cartridge. To screen for potential pretermination products, the purified oligonucleotides were end labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase, resolved by electrophoresis on a 10% polyacrylamide gel, and exposed to Kodak XAR film for 3 h with intensifying screens at −80°C. The positions of the outer and inner sets of primers in the HXB2 isolate have been published previously (80).

PCR amplification. One microgram of PBMC DNA was amplified in 100 μl of reaction mixture containing 3.0 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 20 μM each of the four deoxynucleoside triphosphates (dNTPs), 0.5 μM each of the outer oligonucleotide primer pairs, and 2.5 U of *Taq* DNA polymerase. The reactions were performed in a Perkin-Elmer Cetus automated thermal cycler programmed for 32 cycles of denaturation at 98°C for 10 s, annealing at 50°C for 15 s, and extension at 72°C for 2.0 min. A final extension at 72°C for 10 min was added to the last cycle. A 5-μl aliquot was reamplified in a 100-μl reaction mix containing 0.5 μM each inner oligonucleotide primer pair with the same cycle profile as above. Specific precautions to avoid template and amplified product carryover, including physical separation of processing areas and use of positive-displacement pipettes, were in effect at all times (80). Stringent quality control to prevent PCR contamination was provided to ensure against cross-contamination of product DNA before, during, and after in vitro amplification. Added precautions were taken at the analysis stage when doing cross-comparisons of all sequences included in the study as well as comparisons with common laboratory strains.

Molecular cloning. To screen for the appropriate-sized product, a 20-μl aliquot from the final inner nested amplification was resolved by electrophoresis on a 1.0% agarose gel. The remaining 80 μl of reaction mixture was resolved by electrophoresis on a 1.0% NuSieve GTG low-melting-point agarose gel. The correct DNA fragment was excised from the gel and purified by GeneClean II (Bio 101), phenol-chloroform extracted, and ethanol precipitated. The purified product was eluted in water and digested with the specific restriction endonucleases in the appropriate buffer at 37°C for 16 h. An aliquot of the restricted product was resolved by electrophoresis on a 0.8% agarose gel to screen for any potential restriction sites which would select against a particular variant in the population.

The restricted product was inserted into the multiple cloning site of vector pGEM3zf(−) (Promega), and 0.1 μl of the ligation product was used to transform competent JM109 bacterial cells. Transformed colonies were color selected on ampicillin-isopropylthiogalactopyranoside (IPTG)-5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) agar plates. White colonies were chosen at random and amplified in culture. The bacterial DNA was recovered from a small-scale plasmid preparation. A portion of the plasmid DNA was digested with the appropriate restriction endonuclease, and the insert was screened for appropriate size and relative quantity by electrophoresis on a 0.8% agarose gel.

DNA sequencing. The double-stranded plasmid DNA was sequenced in both forward (M13-21 universal primer) and reverse (M13 reverse) directions with Dye-Deoxy terminators (Applied Biosystems, Inc.). Approximately 1 μg of the double-stranded DNA template was added to 80 mM Tris-HCl (pH 8.3)–2 mM $MgCl_2$–20 mM ammonium sulfate (pH 9.0)–7.5 μM each dNTP–1 μl of each Dye-Deoxy terminator–3.2 pM either oligonucleotide primer–8 U of AmpliTaq DNA polymerase. The total reaction volume of 20 μl was overlaid with white light mineral oil to prevent evaporation. DNA sequencing was performed in an automated thermal cycler programmed for 25 cycles of denaturation at 98°C for 1 s, primer annealing at 50°C for 15 s, and extension at 60°C for 4 min. The sample mixture was then eluted through a spun column (Bio-Rad) and ethanol precipitated to remove unincorporated oligonucleotide primers and dNTPs. The amplified product DNA was added to 5 μl of 5:1 deionized formamide–50 μM EDTA and denatured at 95°C for 3 min. The Dye-Deoxy-labeled sequencing reaction mixes were resolved by electrophoresis on a 6% polyacrylamide gel and analyzed with an ABI 373A automated DNA sequencing system.

Data analysis. Comparisons were done based on various breakdowns of the available sequence data. PIMA (65, 66), an algorithm that incorporates a gap penalty and takes into account amino acid side chain chemistry, was used to generate initial amino acid alignments. Both protein and DNA alignments were then manipulated with the multiple alignment sequence editor (MASE) (19). Simple sequence similarity comparisons were made by using MASE after removing positions in the alignment in which gaps had been inserted to maintain the alignment. These were calculated as Hamming distances, or $(1 - s) \times 100$, where s is the fraction of shared sites in two aligned nucleotide sequences. Hamming distances were appropriate for use in this context because we were not using them to establish phylogenetic relationships, but rather to compare overall genetic diversity in different sets of genetic sequences. To determine the significance of the distinctive distributions of HIV-1 sequence distances, Wilcoxon rank sum statistical tests were run using the SAS statistical package (SAS Institute Inc., Cary, N.C.) Phylogenetic trees (29, 70) were generated by using parsimony (PAUP) (69) and maximum likelihood (PHYLIP) (21, 22).

Signature patterns were defined by comparing the difference in entropy among brain- and blood-derived sequences in each position in an alignment of all sequences from all patients. The entropy $H(i)$ (4) is defined in terms of the probabilities, $P(s_i)$, of the different symbols, s, that appear at a given position i (e.g., in this case, we are considering the symbols $s = A, S, L, \ldots$, for the 20 amino acids Ala, Ser, Leu, . . .). $H(i)$ is defined as:

$$H(i) = - \sum_{s = A,S,L,\ldots} P(s_i) \log P(s_i).$$

Using entropy as a measure of relative conservation takes into account both the variety and frequency of observed amino acids in each position. Potential N-linked glycosylation sites were considered as either unlinked amino acids or as units, where an asparagine (N) of an N-X-(T or S) sequon was distinguished from other asparagines.

To define a signature pattern objectively, we calculated the difference in $H(i)$ between the blood set and the brain set of viral sequences at each position in the amino acid sequence alignment. The signature pattern consists of amino acids in particular positions of the alignment that were very common or perfectly preserved among the brain-derived sequences and variable in the "background" population of viral sequences from the blood. To test the statistical significance of an observed entropy difference, a Monte Carlo randomization was used. The blood and brain viral sequence sets were combined into a large pool and then randomly partitioned into two data sets of the same respective sizes as the original blood and brain data sets by random-with-replacement selection. The entropy differences for each position in 10,000 randomized data sets were calculated for all of the randomized sets to determine a "background level" of entropy difference (39). For a conserved signature site to be considered statistically significant, none of the positions in any of the randomized sets could yield an entropy difference as great as the observed difference in the real data set ($p = 0.0001$). The sign of the entropy difference in the real data indicates whether the observed conservation was in the brain set or in the blood set.

PCR-driven in situ hybridization. Paraffin-embedded tissue sections fixed with a Streck (Omaha, Nebr.) tissue fixative were cut, adhered to silanized slides with 3% (vol/vol) Elmer's glue, and deparaffinized in xylene. After dehydration in graded alcohols and rehydration in phosphate-buffered saline (PBS), the sections were digested in 30 μg of proteinase K per ml in 20 mM Tris-HCl (pH 7.4)–0.5% sodium dodecyl sulfate (SDS) for 1 h at 37°C. The slides were washed twice in PBS for 5 min and dehydrated in graded alcohols (16).

Slides were placed on a Coy Slidecycler (Coy Laboratories), preheated to 82°C, and tissue sections were rehydrated in 40 μl of PCR mixture (10 mM Tris-HCl [pH 8.3]; 50 mM KCl; 1.5 mM $MgCl_2$; 0.25 mM dATP, dCTP, and dGTP; 0.14 mM dTTP; 12.9 μM dUTP-11-digoxigenin; 100 pM each forward and reverse primer; 0.001% [wt/vol] gelatin) preheated to 94°C. *Taq* polymerase (1.0 μl; 5 U of AmpliTaq [Perkin Elmer, Norwalk, Conn.]) was added to one of two sections on each slide. Sections were individually coverslipped, sealed with nail polish, and cycled for 35 cycles (94°C for 1 min, 55°C for 2 min, and 72°C for 2 min). Product DNA was hybridized in situ to specific or nonspecific (CMV IE-3 probe) biotinylated oligonucleotides as described previously (15).

Nucleotide sequence accession numbers. The sequences studied here have GenBank accession numbers U05360 through U05568.

RESULTS

HIV-1 sequences from blood and brain. Figure 1 shows the alignments of the amino acid translations of the nucleotide sequences. The coding potentials of the V3 and V4-V5 regions of env were maintained in most viral sequences. Potential inactivating mutations were observed in both the V3 (five frame shifts and two stop codons) and V4-V5 (two stop codons) regions. There were a total of nine inactivating

A

```
   SENFTNNAKTIIVQLNESVVINCTRPNNNT.RKSIPIGPGRAFYTTGEIIGDIRQ
   ...F...........................S..I....A...T.EI......
Patient 1 BLOOD:
   -D-------I------Q--E-K-I------.--G-A-----TV--AER------- 2
   ---S-----I------K-IE-K-I------.--G-A-----TV--AER-------
   ---------I------K-IE-K-I------.--G-A-----TV--AER-------
   -----T---I------H--E-K-I------.--G-T-----TV--AER-------
   -D--S----I------Q--E-R-I----T-.--G-T-----TV--AER-------
   --I------I------Q--E-K-I------.--G-T-----TV--AEK-------
   -DT--D---I------K-IE-K-I------.--G-A------V--AEG-------
   ---------I------Q--E-K-I------.--G-T-----TV--ADR-------
   ---I-----I------K--E-K-I------.--G-T-----TV---ER-------
   ---------I------S--A-K-I------.--G-A-----TV---ER-------
   -D-S-----I------K--E---I---T--.TQG-N------S----D-------
   ---------I------Q--E-K-I---T--.-NG-A-----TV-SADRL------
   ---S-----I------Q--E---I------.-QG-T-----TV--ADN-------
   ---------I------K-IE-K-I------.----N-------------------
   -D--S----I------Q--E-K-I------.----N-------------------
   ---------I------Q--E-K-I------.----N------V------------
   YN-V-D-T-I------N--G-K-I------.----N-------------------
Patient 1 BRAIN:
   -S-------I------T--E----------.----N------------------- 8
   -S-------I------T-------------.----N------------------- 5
   -S-------I---H--T-------------.----N-------------------
   -SK------I------T----------S--.----N-------------L-----
   -N---D-T-I------S--E----------.----N-------------------
   -S----T--I------T-------------.----N------------------- 2
Patient 3 BLOOD:
   ----SD------------IA----------.--R-TL----V-----Q------R 3
   ----SD------------IA----------.----N--------A--D------- 3
   ----SD------------I-----------.S-R-TL----V-----Q-V----R
   ----SG------------IA----------.--R-TL----V-----Q-V----R
   --KTSD------------I-----------.S-R-TL----V-----Q-V----R
   ----SD------------I-----------.--R-TL----V-----Q-V----R 3
   ---LSD------------IA----------.--R-TL----V-----Q------R
   ----SD------------I---------T-.--R-TL----V-----Q-V----R
   ----SD------------IA----------.--R-TL----V-----Q-V----R 2
   ---LSD---P--------IA----------.--R-TL----V-----Q------R
   ----SD------------IA----------.----N--------A--D------R
   ----SG------------IA----------.----N--------A--D-------
   ----SD------------I-----------.--R-TL---GV-----Q-V----R
   ----SDT-----------I--------I--.--R-TL----V---S-H-V----R
   ----SDT-----------IA-------II-.--R-TL----V---S-Q-V----R
   ----SDT-----------IA-------II-.--R-TL----V---S-H-V----R
   ----SDT-----------IA--------I-.--R-TL----V---S-HLV----R
   ----SD------------I--------II-.$---N--------AS-D------H
   ----SDT-----------IA------.-T-T--R-TL----VS----Q-L----R
   ----SDT-----------I-------.-T-T--R-TL----V---S-HLV----R
   ----SDT-----------I-----------.--R-TL----V---S-H-V----R
   ----SD------------IAL-----.-T-T----N--------AS-D------H
   ----SDT-----------IA-------II-.--R-TL----V---S-HLV----R
   ----SDT-----------IAL------II-.--R-TL----V---S-HLV----R
   ----SDT-----------IA--------I-.--R-TL----V---S-HLV----R
   ----SD------------I---------I-.--R-TL----V---S-H------R
   ----SD------------IAL--S--.-T-T-LR-TL----V---P-H------R
   ----SDS-----------IA----------.--R-TL----V-----H-V----R
   ----SD------------IA----------.----N-----V-----D------R
Patient 3 BRAIN:
   ----SD-V----------I-------S---.-R--N-----------Q-------
   ----SD-V----------I-----------.-------------A---------- 5
   ----SV-V----------I-----------.-------------A----------
   ----SD-V----------I-----------.-------------A------N---
   ----------------A-I-------S---.-R--N-----------Q------- 4
   ----SD-V----------I---------T-.------------HA----------
   ----------------A-I-------S-I-.-R--N---------S-H------H
   ----SDTVT---------I---------T-.-------------A----------
   ----------------A-I-------S---.-R--N--------A----------
   ---------A------A-I-------S---.-R--N-----------Q-------
   ----SD-V----------I-----------.------------DA----V-V-.-
Patient 5 BLOOD:
   ---L-------------T-T-T----S---.--R-A------V---EQ------R
   ---L-------------T-T------S---.--RLA------V---EQ------R
   ---L-------------T-T------S---.--R-A-----TV---EQ------R
   ---L-------------T-T-T-P--S---.--R-A------V---EH-L----R
   ---L-T---I-----T-T-L-T----ST--.----T------------VT-----
   ---L-------L----KT-I-I----S---.----T-----H---R-R.P-----
   ---I-----I------KT-I------S---.----T-------PPP-QLT-----
   ---------I-------T-E---P--S---.----T----G-----E.-L-----
Patient 5 BRAIN:
   ----A----I-------T-E------P---.----T------L---.--------
   ----AI---I-------T-E------T---.----T------L---.--------
   ----A----I-------P-E------T---.----T------L-P-.--------
   ---------I-------T-E------T---.----T------L---.--------
   ---------I-------T-E-T--------.----T------L---.----L-E-
   ---------I-------T-E----------.----T------LS--.--------
   ---------I-------T-E----------.----T------L---.-------- 2
   ------T-NI-------T-E-T----.-T-T--R-T------LSP-.--------
   ---------I-------T-E----------.----T------LF--.--------
   ---------I-------T-E-------T--.----T------LS--.--------
   --------NI-------T-E----------.-----------LF-P.--L-----
```

```
   SENFTNNAKTIIVQLNESVVINCTRPNNNT.RKSIPIGPGRAFYTTGEIIGDIRQ
   ...F...........................S..I....A...T.EI......
Patient 2 BLOOD:
   -D---D---------K-A-E----------.------------------------ 4
   -A---D---------K-A-E-------T--.----------------D-------
   -D---D---------K-A-E-------T--.----------------G-------
   -D-L-D---------KDA-EV---------.------------------------
   -A---D---------K-A-E----------.------R-----------------
   -D---D---------K-A-E-------T--.------------------------
   -D---D---------K-A-E----------.------R--------------X--
   -D---D---------K-A-E----------.----------------D-------
   -D---D---------T-A-E-------T-I.--R-------------G------
   -D-L-D---------K-A-E----------.----------------D------$
   -A---D---------K-A-E-------T-I.--R---------------------
   -D--A----------K-A-E-------T-I.----------------D-------
   -D--A----------K-V-E----------.---V---S---------L----W-
   -D---D---------K-A-E------T---.---V-L----------------W-
   -A---D---------K-A-E----------.---L-V------------V-----
Patient 2 BRAIN:
   TD-H---T-S-----L-A-EV--S--T---.---V-------------V---V--
   -D---D---------K-A-E----------.--------R-------K-------
   -D-------------K-A-E----------.---V-------------------- 2
   -A-------------K-A-E----------.---V--------------------
   -V---DS--------K-A-E----------.----------------D-T-----
   -V---D---------K-A-E----------.---V------------D-------
   -D-------------K-A-E----------.------------------------ 2
   -N--AD---------K-A-E----------.---V--------------------
   -D--A----------K-A-E----------.------------------------
   -A---D---------K-A-E----------.------------------------
Patient 4 BLOOD:
   -----DX--S-----------------T-.-----L-----I-A--H-V-----
   -------V-V---------E----------.-----L-----I-AA-H-V-----
   ---------S--------------------.-----L-----I-A--H-V-----
   -----D---S--I--------------I--.-----L-----I---ET------H
Patient 4 BRAIN:
   ---------S-----------------T-.-----L-----I-A--H-V-N--H
   --P--D---S--------------------.----------------Q------K
   --X--D---S--------------------.-----------S----Q------X
   ---------S-----------------X--.-----L-----I-A--Q-V-N---
   -----D---S--------------------.----------------Q------K 3
   -A----H--S--------------------.-----L-----I-A--Q-V-N---
Patient 6 BLOOD:
   ---L-----------EKA-N-T-E--H---.----I--------A--D-------
   ---------------K-A-N-T-E--H---.----N--------A--D-------
   ------Y--------K-A-N-T-E------.----I--------A--G-------
   ---------------T-A-N-T-E------.----N--------A--A-------
   ---------------R.--N-T-D------.----N--------A-.D-------
   ---------------K-A-N-T-D------.----F--------A--A-------
   -----------L---K-A-N-T-E------.----N--------A--A-------
   ---------------K-A-N-T-E---TT-.----N---R----A--D-------
   ---------------K-A-N-T-E---T--.----I--------A--D-------
   ---------------K-A-N-T-E------.----I--------A-.D-------
   ---------------K-A-D-T-E------.----N-------HAP.D---N---
   ---------------KKA-E-T-E------.----N-------HA-.D-------
   ---------------K-A-N-T-E------.----N-------HA--DV--N---
   ---------------K-A-N-T-E--T---.----N--------A--DV--N---
   ---------------R-A-N-T-E------.----N--------A--DV--N---
Patient 6 BRAIN:
   ---------------Q-A-N-T-E------.----N--------A--D------- (T1)
   ---------------K-A-N-T-E----I-.---FY--------A--D------- (T1)
   --------N------K-A-N-T-E------.----N--------A--D------- (T1)
   ---------------K-A-N-T-E----T-.----N--------A--D------- (T1)
   ---------------R.--D-T-E------.----N--------A-.D---N--- (T1)
   ---------------K-A-N-T-E--T-T-.----T--------A--D------- (T2)
   ---------------K-A-E-T-G------.----N--------A-.D---N--- (T2)
   ---------------K-A-N-T-E------.----N--------A-.D---N--- (T2)
   ---------------Q-.LN-T-E---TT-.----N--------A--D------- (T2)
   ---------------Q-A-N-T-E------.----I--------A--D------- (T2)
   ---------------RK.LN-T-E----T-.----N--------A-.D----L-- (T2)
   ---------------K-A-E-T-E---T--.----I--------A-.D------- (T2)
```

B

```
   MHSFNCGGEFFYCNSTQLFNSTW..N?T......GK?.NTGGNDTITLPCRIKQIINMWQEVGKAMYAPPIRGQIRCSSNITG.LLLTRDGG.NNNNET?.......EIFRPGGGDMR
Patient 1 BLOOD:
   --T----E--P---T-K-----Q..-G-DQNVTERAN.P-E-----------HL-T---R-RQ------------T-----.--------.-SS-T-.......-T---A----K
   T-------------T-K------..DG-DWNVTE-AN.--E-----------H-------R----T----------T-----.--------.-SSS-NET.....-----A----K
   T-------------T-K------..DG-DWNVTE-AN.--E-------------------R--N------------T-D---.--------D.-SSS-NET.....-----A----K
   T------E--L---T-K-----$..DS-D$NVTERAT.--ER----A-----QH------R-RN-------K-----T-----.--------.-E-SSSETEI...-----T--G-K
   T-------------T-K------..-G-DWNVTE-SN.---E----------H------R--E-------------T-----.--------.-SSS-NET.....-----A----K
   T-------------T-K------..DG-DWNVTE-PNRP-E-Y---------T-------R---------K-----T-----.--------.-SSS-NET.....-----A----K
   T----------F--T-D------..DG-DWNVTE-PNRPPE-------------------R---------------T-----.--------.-SS..NET......-T---A----K
Patient 1 BRAIN:
   TN--------S--------T--G..LISTG.STE-SI.S-D-SE-------MMP--D---K----------S-P---------.--------.-S---S.......----------
   -NH-------S---T-K--I--G..-G-DWGVTD-SI.---E---P------NH---I-KR---------------T-----.--------.DD-SSNET.....-----A----K
   --H-----------T-K------..-G-DWSVTE-SN.---E------------------R---------------T-----.--------.DD-SSNET.....-----T----K
   -YH-----------T-K------..-G-DWSVTE-SN.---E------------------R---------------T-----.--------.DD-SSDET.....-----T----K
   --H-----------T-K------..-G-DWSVTE-SN.---E------------------R---------------T-----.--------.DD-SSSET.....-----A----K

Patient 2 BLOOD:
   T------------------G--F.-S-WST..E-SN.--E-S----------F---------------S-----------.--------.----GS.......-----E-----
   T--------------P----G--F.-S-WST..E-SN.--E-S----------F---------------S------T---.--------.----GS.......-----E-----
   T----------------------..YD-KEL.....N.S-.-EM.------------RR-------F---------------VS-------.T-DT--T.......-F------E--
   -----------------------..-D-KEL.....N.S-.-EM.------------R--------------------------.--------.--TDR-T.......----------
Patient 2 BRAIN:
   --N--------------------..HD-KEL.......TST-EM.--------H---------R------------------.--------N------.......P-V---------
   -----------------------..-D-KEL.......-ST-EM.--------H---R------------------------.--------.S-T-G-.......P----------
   --N--------------------..-D-KEL.......-ST-EM.--------H------D--R------------------.--------N------.......P-V---------
   -----------------------..-D-KEL.......-ST-EM.------------R---------------R-------A-.--------.--T-G-.......P----------
   --N--------------------..-D-KEL.......-ST-EM.--------H---------R------------------.--------N------.......P-V---------
   -----------------------..-D-KEL.......-ST-EM.------------R-K----------------------.--------.--A-G-.......P----------
   -NN-------P----P------G..-D-KEL.......IGT-EM.------LT---T--K---R------------------.--------N------.......P-V---------
   -N-------LS----S---T---..-D-KEW.......TRT-EL.--F---LY--LTR-K---N--F---------------.--------.--A-G-.......P----------
   -NN-------P----S---I--G..-D-KEL.......TGT-EM.T-----LN-F-T--K---R--F---------------.--------N------.......P-V---------

Patient 3 BLOOD:
   -YT---------------T---..-IA......-EP.HPTAGEN--------H-----------------K-----------.--------.---K-DNNKT...----------
   --T-------------------..-I-......-EPNIA.-K-----------------------------G-----------.--------.---K-DNNKT...----------
   --T-----------------S-SF-S-W..NSTEESNS-.EG-P------LTHL-S-S-----P-----------V------.--------....-N-TNA.....-T---------
   --T-----------------S-SF-S-W..NIT-EPNI-.-K---------HH-------------P---G-----------.--------.---K-DNNKT...----------
   --T-------------------..-I-......-EPY--.-GEN-----------------------------G-------D--A.--------.-S-K-DNNKT...----------
   --IS-----SL-----------..-ID......-EPY--.-REN-------I------------LS------------Y-A-.-------D....-N-NYA.....-T---------
Patient 3 BRAIN:
   -YT------------S-P-----..-S-......--PNI-.-E---I-------------------------------S---.--------.-----SD......P----------
   -YT-------V-------ST---..-S-......--PNI-.-E---I-----I-L-T------------------------.--------.-D---SD......P----------

Patient 5 BLOOD:
   ----T-----P---T----S---..-E-VI.NQGPYN.T--D-E----Q---------------------K---S-V-T---.--------VKS--T-NSTNTAEV----------
   --------------T----T---..-E-KI.KEGPYN.T--.------Q-------------------------S-V-T---.--------VKS--T-NTTNTTEV----------
   ----T---------T----T---..-EAII.NQGAIY.HYRKLKQSSS---------------------K-P-S-L-----.--------E-TD-A-..DNNT..-V---------
   ----------P---T----T---..-E-II.TQGPYN.T--N-E-L--QS--T-L-T------------------S-L-Y---.--------.R-ADS-.TDNNT..----------
   ----T-----L---T---------..SE-IT.KQGPYN.T--N-E----Q---T-----------------K-----E-----.--------..-D-T-..DNNT..----------
   ----T-----P---T----T---..-E-II.KQGPYN.T--N-E-L--Q---T-L----------------T-----V-----.--------..-D-TN..DNNT..----------
   ----T-----LC-----------..-G-....EGPYN.T--.-E---------------------------K-----L-H---.--------..H-PYQ...YHR..----------
   ----------SF--T-H--T---NE-V-....EESYE.T--.-E---------------------------K-----E-----.--------..-D-T-..DNNT..----------
Patient 5 BRAIN:
   -----R----L--D---------NE-V-....EESYE.T--.-E---------------------------K---S-L-----.--------.------...NPT..-T---------
   -------------D---------NE-V-....EESYE.T--.-E---------------------------K---S-L-----.--------.------...TTT..-T---------
   -------------D---------NE-V-....EESYE.T--.-E---------------------------K---S-L-----.--------.------...NTT..-T--------- 2

Patient 6 BLOOD:
   T----------F--T--------YS-N-W.....EG.ERLS----I-------L-----S----------K-R---------.PVANKKMVG.---N-.DGNES..----------
   T----------F--T--------YS-N-W.....EG.ERLS----I-------L-----S----------K-R---------.--------.---N-.DGNES..----------
   T---I------F--T--------YS-T-.......RR.GE-LRTN-------P-V-----S---L-----K-R---------.--------.---N-.DGSES..----------
   T---------SS--TS-------YS-N-.......RK.GE-LSTN-------P---H---S----------K-R---------.--------.---N-.DGNRI..-T---------
   T----------F--T----T---YS-N-.......RR.GE-LRNN---------V-----S----------K-R---------.--------.---N-.DGNES..----------
   T----------F--T--------YS-N-.......RR.GE-LSNN---------V-----S----------K-R-----T---.--------.---N-.DGSQS..----------
   T----------F--T-K------NNTGE.......SK.G.IS----I-------L-----S------------L-T-------.--------.---N-.DGNRI..-T---------
   T---T------S--TSP--T---YK-S-W.NNT.-EG.ERLC----I-----T-V-----S---L-----K-RL--L-----.TVGYQLD-.G---N-.DGNRI..-T---------
Patient 6 BRAIN:
   T----------F--T--------YC-N-......WEG.ERLS-N.T--------V-----S----------K-R-----H---.--------....-I-.TRSEI..----------- (T1)
   T---------PF--T-K------..-N-......-ES.EGLS-N.---------V-----S----------T-R--------A.-S--------....-I-NRS.EI..----------T (T1)
   T---------SF--T--------YS-N-......RNG.EGLS-N.---------L-----S-----F---K-R-----H---.--------....-I-TRS.EI..----------- (T1)
   T---------SF--T-K------..-N-......-ES.EGLS-N.---------V-----S-----F---K-R-----H---.--------....-I-ARS.EI..----------- (T1)
   T---------SF--T-K------..-N-......-ES.EGLS-N.---------V-----S----------K-R---------.--------....-I-NRS.EI..----------- (T2)
   T---T------F--T----D---YS-N-......R-G.EGSS-N.V--------V-----S----------K-R---L-----.--------....-I-NGS.EI..-T--------- (T2)
   T----------F--T----D---YS-N-......R-G.EGLS-N.---------L-----S----------K-R-----H---.--------.---T-.DGNES..----------- (T2)
   T---------LF--T-K------..-N-......-EI.EGIS----I-------L-----S----------K-----------.--------..-ITNTNGNES..----------- (T2)
   T----------F--T--------YS-N-......R-G.EGLS-N.---------L-----S----------K-R-----T---.--------.---T-.DGNES..----------- (T2)
   T----------F--T--------YS-N-......W-G.ERLS-N.T--------V-H---S----------K-R---------.-S--------....-I-NRS.EI..----------- (T2)
   T---------LF--T--------YS-N-......W-G.EGLR-N.T------T-V-----S----------K-R---------.-S--------....-I-NGS.EI..-T--------- (T2)
   T----------F--T----D---YS-N-......R-G.EGLS-N.---------L-----S----------K-R---------.--------.---T-.DGNES..----------- (T2)
```

FIG. 1. Alignment of the V3 (A) and V4-V5 (B) region HIV-1 envelope amino acid sequences obtained from brain biopsy or blood samples. Viral sequences are aligned to the consensus sequence of the brain sequences from the six individuals, shown at the top of each column. Directly underneath the complete consensus in panel A are the brain sequence signature amino acids as defined in Table 2 and in the text. Amino acids are indicated in the one-letter code. Amino acids in the alignment that share identity with the brain consensus sequence shown at the top are indicated by a dash. If two or more clones from a given set are completely identical over the region sequenced, only one is shown; the numbers at the end of a sequence indicate the number of clones represented by that sequence. A dollar sign symbol ($) indicates a stop codon, and an X indicates a frame shift. A period indicates a space inserted to maintain the alignment. Brain biopsy samples were taken from patient 6 at two time points, indicated by T1 or T2 at the end of the sequences; T1 corresponds to the time that the blood sample was obtained, and T2 was 6 weeks later.

TABLE 2. Frequency of signature amino acids[a]

Sequence derivation (no. of sequences)	Frequency of occurrence								
	F	S	I	A	T	E/D	I	NCT	NNT
Blood									
Patient 1 (18)	0.72	0.22	1.00	0.33	0.39	0.28	0.94	0.11	0.89
Patient 2 (18)	0.89	0.89	0.89	1.00	1.00	0.89	0.94	1.00	0.67
Patient 3 (36)	0.92	0.22	0.22	0.19	0.63	0.22	0.86	1.00	0.83
Patient 4 (4)	1.00	1.00	0	1.00	0.75	0	1.00	1.00	0.75
Patient 5 (8)	0.13	0.50	1.00	0.75	0.75	0.25	0.63	0.38	0.86
Patient 6 (15)	0.93	1.00	1.00	1.00	0.93	0.80	0.80	0	0.87
Brain									
Patient 1 (18)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.94
Patient 2 (12)	0.92	1.00	1.00	1.00	1.00	0.92	0.92	1.00	1.00
Patient 3 (18)	1.00	1.00	1.00	1.00	0.94	0.61	1.00	1.00	1.00
Patient 4 (8)	1.00	1.00	0.63	1.00	1.00	0	1.00	1.00	1.00
Patient 5 (12)	1.00	0.92	1.00	1.00	0.92	1.00	1.00	0.83	0.92
Patient 6 (12)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0	0.83

[a] The conserved amino acids of the envelope V3 region signature pattern are shown at the top; these amino acids correspond to the signature sites shown in Fig. 1A. The number of sequences of each sample is shown on the left, following the patient number. The frequency of the occurrence of the signature amino acid in each sample is given. Glutamic acid (E) is the sixth signature amino acid. Aspartic acid (D) was also very common and a conservative substitution, and the statistical significance of the conservation of that site was observed when signature analysis was done using the individual amino acids or a modification of the input which grouped D and E as a single character. Therefore, D or E is allowed for the frequency counts. The frequency of the glycosylation sites in the V3 loop is also given. The V3 glycosylation sequons were more conserved among brain sequences, in all 10,000 of the randomizations; however, the conservation was critically dependent on the inclusion of patient 1 for the first sequon (NCT) and patient 2 for the second sequon (NNT), and because of this dependence on a single patient, the conservation was deemed less reliable than that of the other signature pattern elements.

mutations in 54,093 bases sequenced, six found in the 28,872 nucleotides of blood-derived viral sequences and three found in 25,221 nucleotides of brain-derived viral sequences. The four cysteines encompassed in the regions sequenced that are involved in disulfide bonding and in the formation of the V3 and V4 loops (47) were highly conserved, but not perfectly, with two sequences having substitutions in these positions. Such substitutions are also likely to represent inactivating substitutions, as mutations in these positions can have drastic structural consequences for the envelope protein.

Length polymorphism was common in the V4 and V5 region, even when considering only intrapatient sequence sets. In these highly polymorphic regions, duplications or deletions of N-linked glycosylation sites, as defined by an N-X-T or N-X-S sequon, were particularly common among the blood-derived sequences (Fig. 1B). The two glycosylation sequons in the V3 loop tended to be more preserved in the brain-derived than in the blood-derived viral sequences. Because of the biological importance of glycosylation sites (46, 51, 57) and the conservation of these sites among sequences in the Human Retroviruses and AIDS database (56), the information summarizing the variation of these sites is included in Table 2.

Intrapatient sequence variation. A general question of interest in terms of assessing intrapatient HIV-1 variation is whether blood samples are representative of the variants within an individual or if certain forms may be sequestered at inaccessible sites (14). The V3 region diversity within a patient increased when both blood and brain sequences were considered versus only blood sequences (Fig. 2). The median distance of all pairwise intrapatient blood × blood sequence comparisons for the six patients was 4.9%; the same brain × brain median distance was 3.5%, and the same intrapatient blood × brain median distance was 7.7% (Fig. 2). There was no change in observed diversity for the V4-C3-V5 regions (median distances: blood × blood, 7.2%; blood × brain, 6.0%; and brain × brain, 4.4%). The brain-derived sequences were relatively conserved compared with the blood-derived sequences in both regions.

There was an expansive range of intrapatient HIV-1 sequence diversity within this set of HIV-1-infected patients. In the V3 region, the range of intrapatient nucleotide sequence distances was 0.0 to 18.9%, with a median value of 6.3% and

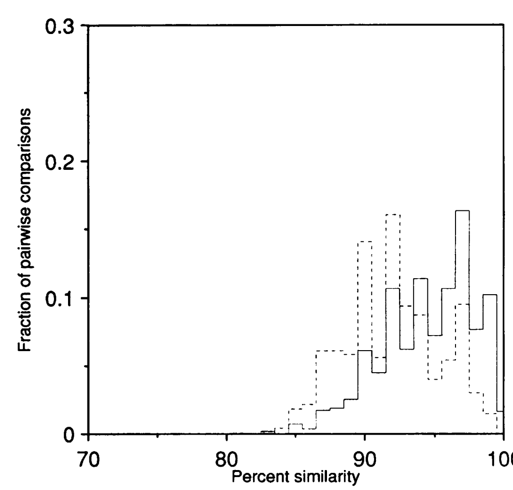

FIG. 2. V3 region intrapatient similarity distributions. All intrapatient similarity comparisons were made for the six study participants. The region compared included 143 bases. The percent similarity was rounded off to the nearest percentage, and the number on the y axis represents the fraction of the total number of pairwise comparisons that had a given percent similarity. For example, 1.6% of the sequences were 100% similar (identical) over the region sequenced. The solid line represents the distribution of similarity scores obtained from intrapatient blood-derived sequences. The dashed line represents the similarity scores of all intrapatient comparisons of blood-derived sequences with brain-derived sequences. Note the shift towards greater diversity when comparing blood-derived (median distance = 4.9%) and brain-derived (median distance = 7.7%) viral sequences rather than blood-derived sequences alone.

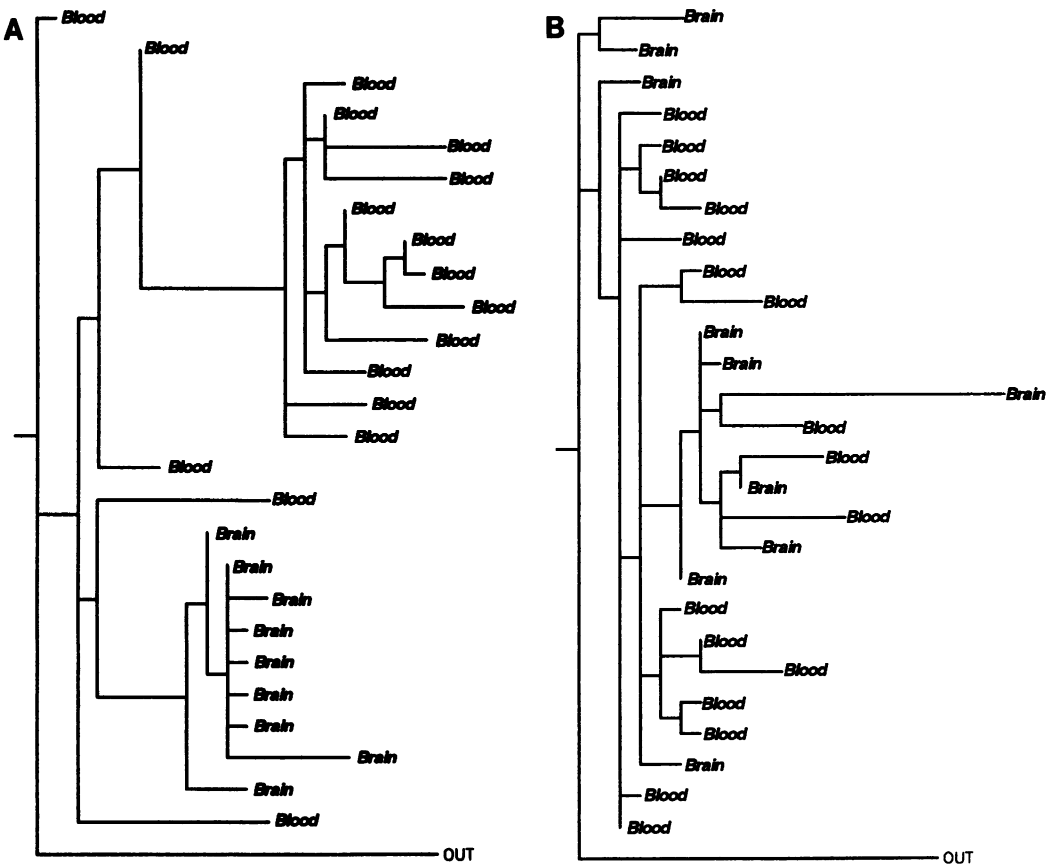

FIG. 3. Phylogenetic analysis of intrapatient sequence sets. Viral sequences from the V3 region from (A) patient 1 and (B) patient 2. The trees shown have the maximum likelihood after three randomizations of the input order. Identical nucleotide sequences were excluded. "Blood" indicates that the taxa are derived from the blood; "Brain" indicates that the taxa are derived from the brain. A sequence from the blood of patient 6 was used as the outgroup sequence for the trees of both patient 1 and patient 2. For patient 1, there were 27 taxa and 163 sites, with 97 varied sites; for patient 2, there were 28 taxa and 158 sites, with 86 varied sites. The trees shown display patterns similar to those obtained by using parsimony and were supported by bootstrap analysis of parsimony trees (see text).

interquartile range of 3.5 to 9.1%. The range of intrapatient nucleotide sequence differences between blood- and brain-derived isolates in the V4 and V5 regions was 0.0 to 17.3%, with a median of 5.2% and interquartile range of 3.6 to 7.2%.

Phylogenetic analyses of intrapatient sequence sets. Two patterns were observed in the phylogenetic reconstructions of genetic lineages of sequences from intrapatient sequence sets. In four of the six data sets, the brain-derived viral sequences were tightly clustered relative to the blood-derived viral sequences (the strongest cases for tissue-specific compartmentalization of virus can be made for patients 1, 5, and 3; patient 4, with much fewer data, seemed to also fall into this category, but less definitively). In the other two data sets (patients 2 and 6), the blood- and brain-derived viral sequences were intermingled. These observations were consistent when two methods were used to construct phylogenetic trees: maximum likelihood (21, 22) and parsimony (69). The V3 region maximum-likelihood trees obtained for patients 1 and 2 were selected to illustrate the two patterns described above (Fig. 3).

The phylogenetic analysis for patient 1 shows the brain sequences forming their own distinct clade, while in patient 2, the brain sequences were intermingled with blood sequences. For the five patients with the V4-V5 region sequence information (all but patient 4), the V4-V5 phylogenetic patterns agreed with the observed V3 region pattern, although fewer sequences were compared. A bootstrap test applied to the parsimony trees (20, 22, 30) generally supported the brain-derived viral sequence clustering patterns obtained by using parsimony for patients 1, 3, and 5. The bootstrap proportions for the reconstruction of the brain-clustering patterns in 100 resamplings of the data were 74 of 100 for V3 and 98 of 100 for V4-V5 for patient 1; 58 of 100 for V3 and 98 of 100 for V4-V5 for patient 3; and 78 of 100 for V3 and 92 of 100 for V4-V5 for patient 6. The bootstrap proportions were shown by Hillis and Bull (30) to be a conservative estimate of the accuracy of reconstructions of true phylogenetic branching order under a range conditions. The relative conservation in length of the V4 and V5 region brain-derived viral sequences from patients 1, 3, and 5 and the length polymorphism in the brain-derived viral sequences of patient 2 and patient 6 also supported the two distinct phylogenetic patterns (Fig. 1B).

Localization of virus in brain tissue by PCR-driven in situ hybridization. Histological examination showed no evidence of characteristic multinucleated giant cells or microglial nodules in any of the six patients. Two of the patients, representatives of the two distinct intrapatient phylogenetic patterns, were selected for further experiments to examine the cell type associated with viral infection. Patient 3 represents an example of a patient whose brain-derived sequences formed a distinct clade. Patient 6 represents an example of a patient whose

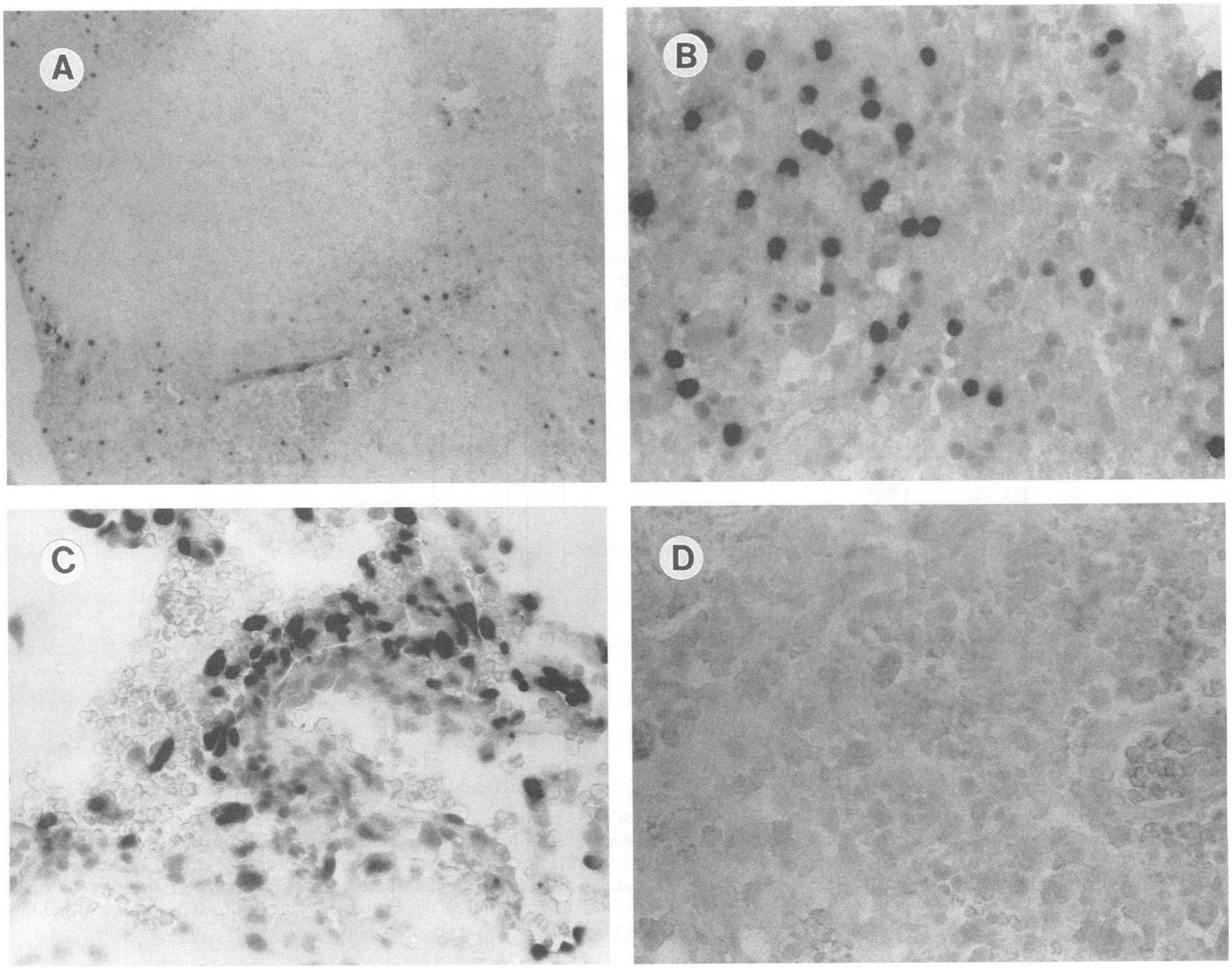

FIG. 4. Detection of HIV-1 DNA in brain biopsy samples by tissue-based PCR-driven in situ hybridization. (A) Biopsy from patient 6, with lymphocytes containing HIV-1 DNA surrounding an area of necrosis. (B) Higher magnification of sample in panel A, showing predominantly lymphocytes containing HIV-1 DNA. (C) Biopsy from patient 3, with perivascular macrophages and endothelial cells containing HIV-1 DNA. (D) Biopsy from patient 6 amplified with HIV-1 *gag* primers and probed with a biotinylated CMV-specific probe (negative control). Magnification: panel A, ×79; panel B, ×316; panel C, ×316; panel D, ×158.

blood- and brain-derived sequences were intermingled in terms of phylogenetic analysis. Interestingly, the histological examination of the infected cells from the brain tissue from patient 3 revealed a strict localization of HIV-1 DNA within perivascular mononuclear cells and occasional microglia (Fig. 4A), and a predominant localization of HIV-1 DNA in lymphocytes within an area of necrotizing encephalitis in patient 6 (Fig. 4B). No HIV-1-infected neurons were seen. These patterns of infection are consistent with our interpretation of the respective phylogenetic analyses of patients 3 and 6.

Interpatient analysis of sequence sets. In the V3 region, all pairwise interpatient nucleotide sequence distances between brain-derived samples (the median brain × brain interpatient distance is 11.8%) were significantly less than the interpatient distances between blood samples (the median blood × blood interpatient distance is 15.2%). For each patient, the blood-derived consensus sequence was compared with all blood-derived sequences from other patients, and the brain-derived consensus sequence was compared with all brain-derived sequences from other patients. The interpatient brain-derived viral sequence distances were closer in each case ($p < 0.0001$) for all patients except patient 4, who still was significantly closer in brain-derived viral sequences than in blood-derived sequences compared with the other patients ($p = 0.007$). Despite the greater conservation of brain-derived viral consensus sequences than blood-derived sequences between patients, this effect is small compared with overall interpatient distance. Phylogenetic analysis showed that blood- or brain-derived viral sequences from each patient cluster together relative to those from the other patients and a background set of North American B subtype sequences (56), despite the greater level of conservation between brain-derived viral sequences than between blood-derived sequences (data not shown). Intrapatient clustering of consensus sequences for all patients was found in 100 of 100 bootstrap replicates of parsimony trees.

In the V4-V5 region, interpatient distances between brain samples (the median brain × brain distance is 14.1%) were equivalent to the distances between blood samples (the median

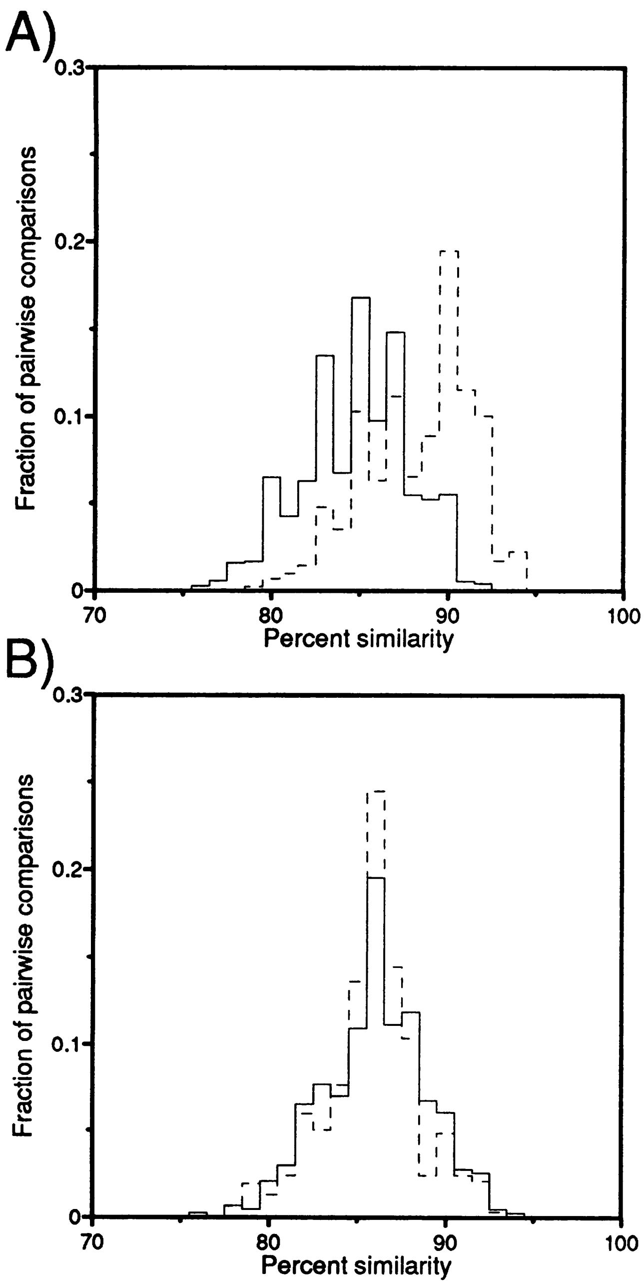

FIG. 5. Interpatient similarity distributions. (A) Distributions of interpatient sequence similarities for V3 region samples (over the same 143 bases used for Fig. 3); (B) distribution for V4-V5 region samples (over 251 bases). As in Fig. 2, the percent similarity was rounded off to the nearest percentage, and the number on the *y* axis represents the fraction of the total number of pairwise comparisons that had a given percent similarity. The interpatient distances for blood samples are very similar in the two regions, with medians of 14.2% in V3 and 13.9% in V4-V5 (B). The distribution of interpatient comparisons in brain-derived sequences of the V3 region is significantly offset, however, with median values of 11.8% in interpatient comparisons of brain V3 sequences and 14.1% in brain V4-V5 sequences (A).

blood × blood distance is 13.9%). This was in contrast to the relative conservation of the brain-derived viral sequences from the V3 region. Figure 5 shows the interpatient similarity distributions for the V3 and V4-V5 regions expressed as histograms. The relative conservation of the V3 sequences is apparent, as the distribution of similarity scores is distinctly closer for the brain sequence comparisons. These results suggest either a convergence to or preservation of a favored form in V3 region brain sequences.

V3 region signature sequence pattern of noncontiguous amino acids. Because of the relative conservation of interpatient V3 region brain-derived viral DNA sequences observed between the six patients, we explored the possibility of identifying signature amino acids that were conserved among the brain-derived viral sequences of the six patients but variable in the blood-derived viral sequences. A simple signature (35, 37) defined by a shift in the most common amino acid in a given position, *i*, in an alignment of blood sequences relative to brain was not ideal for this purpose. We expected situations to arise where both the brain set and the blood set of sequences have the most common amino acid at position *i* but this position is more variable in the blood- than in the brain-derived viral sequences. Therefore, we calculated a formal measure of variability at each position *i* in their combined alignment: the Shannon entropy, $H(i)$ (see Materials and Methods) (4). The statistical significance in observed entropy differences was calculated by using a Monte Carlo randomization method. By using this technique, signature sequence patterns of noncontiguous amino acids were identical for the brain-derived sequences (Fig. 1).

Because a relatively conserved brain-derived HIV-1 sequence set from a single individual may be dominating the signature pattern, the V3 region brain signature pattern may have been an artifact due to brain sequences evolving from a single lineage in one of the patients. To exclude this possibility, patients were excluded from the analysis one by one, and six reanalyses were done with data for only five patients; this ensured that signature pattern sites were not dependent on any single patient's sequences. Signature sites from the original analysis using all six patients were included only if they were selected in all six of the reanalyses. Table 2 shows the frequency of signature amino acids in the blood and brain sets from each patient. Some patients show perfect or nearly perfect conservation of a given amino acid signature site in both blood- and brain-derived sequences; these patient sets are noninformative for these particular signature sites. Intact or close to intact signature patterns are found in most of the brain sequences (Fig. 6). In the blood sequence sets, two classes of viral sequences are observed: those that retain the brain signature, and those that do not. This indicates that the set of viruses that persists in the brain is a subset of the spectrum of viruses found in the blood. No significantly conserved amino acid positions by the criteria described above were observed in the blood-derived V3 sequences or in the V4-V5 region sequences from either blood or brain.

Net charge in the V3 loop: brain- compared with blood-derived viral sequences. The net charge on two protein regions was calculated for all sequences in this study. We considered both the entire V3 loop and a 15-amino-acid fragment of the loop that includes the 4 amino acids at the tip of the loop (GPGR or its analog), 4 amino acids on the N-terminal side of the tip, and 7 amino acids on the C-terminal side of the tip (corresponding to SIPIGPGRAFYTTGE from the brain consensus sequence shown in Fig. 1A). The interior region was used as well as the intact V3 loop because it encompasses elements that have been shown to be critically important for viral phenotype through mutational analysis and variation studies (11, 13, 23, 33, 64, 76, 77). It includes the sites which have been shown to covary and thus may be interactive in terms of structure and function (39). Also, the 15-amino-acid

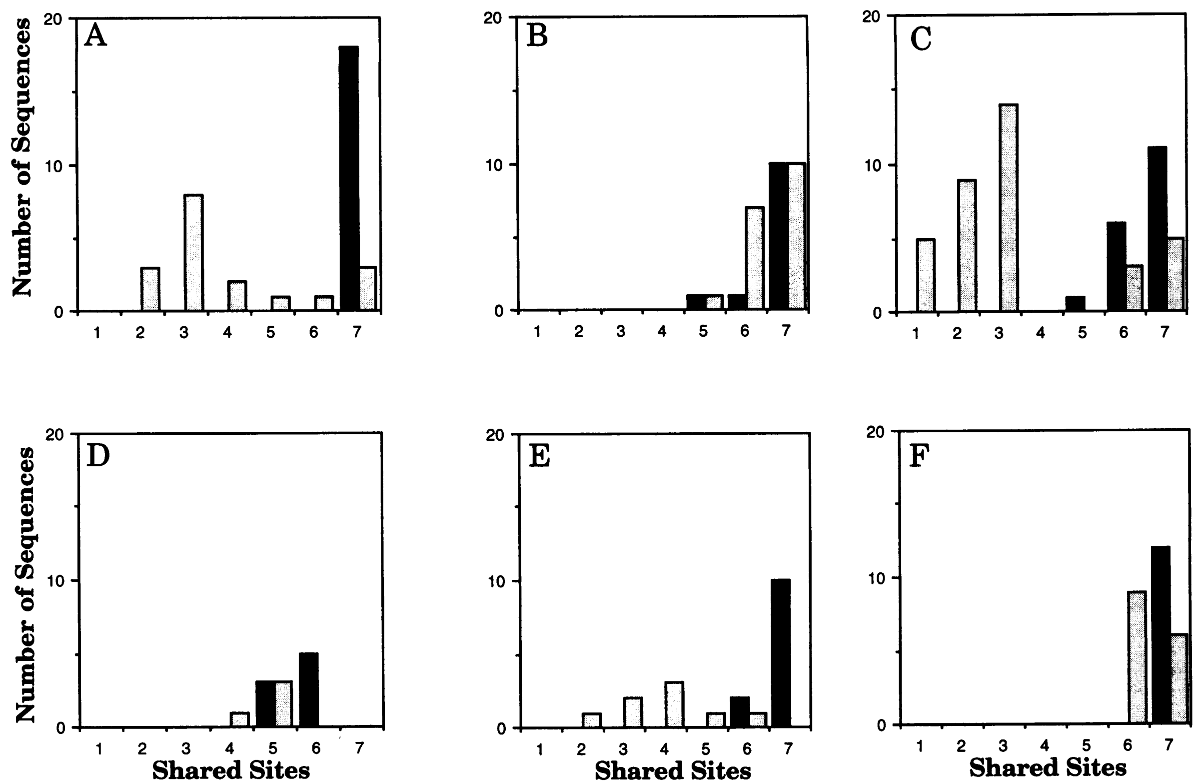

FIG. 6. Distribution of blood- and brain-derived sequences that share a given fraction of the seven-amino-acid brain signature pattern. Solid bars represent brain-derived viral sequences, and shaded bars represent blood-derived sequences. The number of sequences that share a given number of amino acids with the brain signature pattern are indicated. Panels A, B, C, D, E, and F contain data from patients 1, 2, 3, 4, 5, and 6, respectively.

tip gave a clearer distinction than the intact V3 loop when we compared the net charge on the set of syncytium-inducing viral sequences with that of non-syncytium-inducing viruses published by Fouchier et al. (23) (data not shown). The distribution of net charge among blood-derived sequences and brain-derived sequences is shown in Fig. 7. As with the syncytium-inducing characteristics in the Fouchier data set, the internal region was better than the intact loop for distinguishing blood and brain sequence sets on the basis of net charge.

DISCUSSION

Like other retroviruses, HIV-1 is characterized by a high degree of genetic variability. The spectrum of genetically diverse variants found in a viral population within an infected individual is responsive to the selective forces of evolution. In primary infection, there is a relatively narrow distribution of genetic variants (82, 83). The complex mixture of genetic variants that subsequently arise during the course of an infection (31, 38, 42, 73) results from competition and selection among the variants in response to immunologic pressure for change and alterations in cell tropism and replication efficiency (12, 32). Selection due to cell tropism and replication efficiency could result in discrete populations of viruses with distinct genetic and biologic features found within specific tissue types (8, 10, 18, 35, 68).

In this study, we observed a monophyletic population of brain-derived relative to blood-derived viral sequences in four of six patients. Phylogenetic tree analysis of intrapatient genetic lineages for patients 1, 3, 4, and 5 shows a distinct clade consisting of the brain-derived viral sequences, consistent with compartmentalization of viral sequences between blood and brain tissues. These results confirm and extend observations made by Keys et al. (35), where 8 of 10 intrapatient sequence sets tended to show clustering of cerebrospinal fluid-derived viral isolates versus blood-derived isolates (35). Further, the results of our PCR-driven in situ hybridization studies for patients 3 and 6 were consistent with our interpretation of the genetic analyses. The use of viable brain biopsy material, direct amplification of DNA without antecedent passage in culture, and analysis of up to 36 sequences generated from a single patient in our data set allowed a more robust characterization of viral sequence relationships than previous studies (18, 35). Viral compartmentalization in the brain of these patients may have resulted from biological selection, the kinetics of viral replication, or the evolutionary dynamics of infection or because the brain is an immunologically privileged site. The blood-brain barrier, composed of specialized capillary endothelial cells joined by highly restrictive tight junctions (27), could normally inhibit the transfer of virus but, at some point during the infection, allow an exclusive lineage through to establish the infection of the brain tissue. Alternatively, restricted cellular tropism of viruses found in the brain may only allow growth of a selected lineage, resulting in a distinct microenvironment in the brain.

In contrast to the distinct lineages of brain-derived viral sequences observed in patients 1, 3, 4, and 5, an admixture of blood- and brain-derived sequences was observed in patients 2 and 6. The results of histological examinations of the virally infected cell types from the brain biopsy specimens paralleled

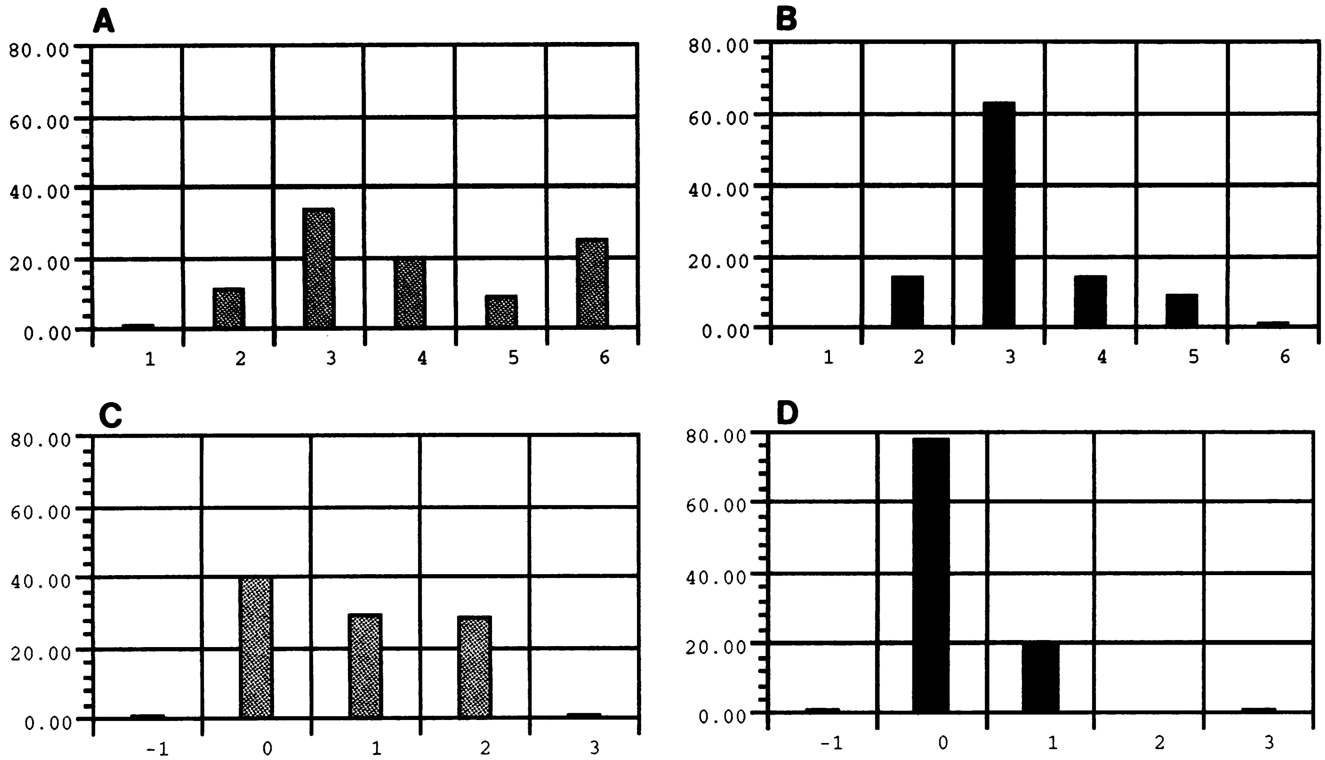

FIG. 7. Net charge of V3 loop sequences in the blood and in the brain. (A and B) Net charge distribution among blood (A) and brain (B) sequences for intact V3 loop sequences. (C and D) Net charge distribution for the 15 amino acids at the tip of the V3 loop, as described in the text. Solid bars represent brain-derived viral sequences, and shaded bars represent blood-derived sequences. The height of the bar indicates the percentage of the total number of sequences that had a particular charge, and the position on the x axis indicates the charge.

those of the phylogenetic analyses. A typical infection of macrophages and microglia was observed in patient 3, who had a monophyletic set of brain-derived sequences. Infiltrating lymphocytes within an area of necrotizing encephalitis, however, were observed in the brain biopsy specimen from patient 6. The infection of the CNS by a virus adapted to brain tissue could account for the neurological signs and symptoms, neuropathology, and the population of distinct viral genetic variants observed in patient 3. The pathology associated with lymphocytes infiltrating the brain could account for the influx of viral genotypes from the blood observed in the brain biopsy sample of patient 6. For patients 2 and 6, disruption of the blood-brain barrier could have occurred as a consequence of the underlying disease process or through blood contamination of the brain sample at the time of biopsy. Through using biopsy- rather than autopsy-obtained brain tissue, however, we have attempted to minimize the potential for blood contamination. Autopsy brain tissue samples, unless taken immediately upon death, may be influenced by postmortem changes that can compromise the integrity of the blood-brain barrier.

To determine the potential for evolutionary dynamics of infection, we examined two consecutive brain samples taken with a 6-week interval from patient 6. A blood sample was taken coincident with the first brain biopsy sample. Phylogenetic reconstructions indicated that the brain sequences from the second time point were intermingled with the blood samples from the first time point rather than the brain sequences from the first time point (data not shown). This information is a further indication that the distinction between the blood and brain tissue was compromised and that viruses were passing between the two compartments in this patient.

In the V3 region of the env gene, the brain-derived sequences were more similar than the blood-derived sequences when comparing differences among patients. This genetic difference was not apparent in the V4-V5 region. These data suggest either convergence or a lack of divergence in the V3 region among the HIV-1 sequences derived from the brain. To search for an underlying commonality among the brain-derived viral sequences, we calculated and compared the entropy of each amino acid position in the blood and brain viral sequence alignments. Seven noncontiguous amino acids were conserved among the brain-derived viral strains. Six of the signature sites were in the V3 loop, and the seventh was in the C2 region. This brain signature pattern was compared with sequences in the Human Retroviruses and AIDS database and found to be highly conserved among sequences from laboratory-adapted macrophage-tropic virus strains (23, 26, 50, 53, 56, 58, 59, 76). Some of the positions in the brain viral signature had previously been identified as important determinants for macrophage tropism, while other signature positions of potential importance had not been identified previously. The signature site in the C2 region was proximal to an inactivating mutation in SF13 studied by Stamatatos et al. (67). This inactivating mutation was able to be partially compensated for by a substitution in the V3 loop (67), thus biologically linking the immediate vicinity of the C2 region signature site to the V3 loop. Additionally, the amino acid substitutions in the tip of the V3 loop of the brain-derived viral sequences had a tendency towards less positive charge than the blood-derived viral sequences (Fig. 7). This charge difference was also apparent when Fouchier et al. examined laboratory-adapted viral strains with syncytium-inducing and non-syncytium-inducing phenotypes and differences in cellular tropism (23).

The signature pattern derived from our viral sequence data set was compared with published CNS- versus non-CNS-derived viral sequences (18, 35, 50). Our signature was highly conserved in the viral sequences from brain autopsy tissue obtained from a single patient by Li et al. (50). Important differences in sample source make our study fundamentally distinct from those of Epstein et al. (18) and Keys et al. (35). These differences include autopsy versus biopsy, cultured virus versus direct PCR amplification from tissue samples, clinical stage and age of the patients, differences in source (cerebrospinal fluid versus brain samples), and viral genetic subtype (18, 35). The signature pattern was conserved in viral sequences from both blood and cerebrospinal fluid obtained by Keys et al. (35), with sequences retaining between five and seven of seven of the signature pattern sites. The conservation found in sequences from this set of blood-derived viruses may have been a consequence of earlier disease stage of many of their study subjects of passage of virus in culture prior to PCR amplification (35). Because of the integrity of the blood-brain barrier, the virus in the cerebrospinal fluid is likely not to reflect the virus that resides in the brain parenchyma. Among the limited viral sequences presented in Epstein et al. (18), our signature was not particularly well conserved in brain sequences relative to spleen, but this may be due to the fact that this study examined viral sequences derived from autopsy material from pediatric patients (18). Also, two of the three patients presented in Epstein et al. had indistinct phylogenetic patterns for blood and brain sequences. This situation was similar to that of patients 2 and 6 in our study, in whom viruses may be trafficking between the two tissues, making it less likely for a signature pattern to be distinctive and apparent.

The hypothesis that the brain signature pattern is conserved due to constraints imposed by cell tropism for monocyte-macrophages or microglial cells is consistent with the growth characteristics of virus isolated from additional brain biopsy specimens. The brain biopsy-derived viral isolates were non-syncytium-inducing in mixed-donor PBMCs and replicated to high titer in cultured monocyte-derived macrophages (16a) as well as in primary adult human brain cultures (16b). Others have also observed similar macrophage-monocyte growth characteristics for brain-derived viral samples (10, 35, 41, 50, 63). The presence of the brain signature pattern in some of the blood-derived sequences (Fig. 6) could reflect the distribution of blood-derived viruses found in PBMCs into cellular fractions, T cells, and macrophage-monocyte cells.

While the occurrence of a brain signature pattern might be due to convergence, a more plausible explanation is simply a lack of divergence and the preservation of certain key elements in protein sequences. Convergence would require that the set of conserved amino acids reemerge independently in the six patients. Evidence from several studies has indicated that there can be selection upon transmission (43, 80, 82, 83) and that the homogeneous form of the virus found before and at seroconversion tends to have sequence elements in common with macrophage-monocyte-tropic viral forms (82, 83) and that these forms tend to be relatively conserved compared with T-cell-tropic forms (54). As macrophage tropism may be a phenotypic characteristic of virus that becomes established in the brain, the signature pattern may not have reemerged in brain-derived viruses independently in the six patients; rather, the brain-derived viruses may retain a pattern commonly selected at transmission. To examine this hypothesis, the brain signature pattern was compared with HIV-1 sequences derived from recent seroconverters (44, 79, 82, 83). It was found to be quite well preserved in early seroconverters (except in one case of a parenteral transmission route [79]). In the study by Wolfs et al. (79), sequences from both donors and recipients from transmission pairs were available, and the signature was well preserved in the early seroconverter recipients (six or seven of seven sites retained) but variable in the donors (most donor sequences sharing only three or four of seven signature sites).

Only one of the signature amino acids was perfectly conserved in the brain sequences; the others showed some variability. Thus, it is important to emphasize that signature amino acids are not strictly required; rather, they tend to be conserved in brain samples from these six patients. Similar conclusions have been reached by comparing conserved amino acids in key positions that are associated with macrophage-monocyte tropism in cell culture (11). Therefore, the brain signature pattern based on our data set should be used only as a guide to highlight positions that may be of biological importance; the statistical significance of the relative conservation of these sites is supported by our analyses. Table 2 can be used as a qualitative guide to assess the relative merit of each signature position. The signature pattern method is generally applicable and may be useful for comparing other HIV-1 sequence data sets, such as mother-infant vertical transmission pairs, samples from long-term immunological nonprogressors, virus from people who progress rapidly to disease, or viruses with different phenotypes in culture.

In conclusion, our data demonstrate that viral sequence heterogeneity exists in different body compartments within an infected individual. Specific charge or conformation changes in the V3 loop may contribute to the observed differences in cell tropism among the brain isolates. Within the brain, microglial cells and multinucleated cells of the monocyte-macrophage lineage are the predominant infected cell types. While neuropathologies commonly arise in HIV-1-infected individuals as a consequence of HIV-1 infection or opportunistic infections, other infected individuals are essentially spared the more devastating aspects of CNS infection. The difference in clinical outcome may depend in some cases upon the host's immune response to the viruses which are able to infiltrate the CNS. Macrophage-tropic forms derived from the CNS are less susceptible to serum neutralization than T-cell-tropic forms (8). The observation that amino acids in positions which are known to be important for antigenic specificity are relatively conserved among brain-derived sequences indicates that immunotherapeutic agents against the forms commonly found infiltrating the brain may be beneficial in terms of preventing or restricting neurological disorders, by augmenting recognition of these forms by the host immune surveillance mechanism.

ACKNOWLEDGMENTS

We thank Gerald Myers, James Theiler, and Catherine Macken for their thoughtful advice concerning analytical methods and Kersti MacInnes and Patricia Otto for technical support. We also thank Carla Kuiken for her suggestions for the manuscript.

This work was supported by NIH NIAID-DOE interagency agreement 3-Y01-A1-70001-11 (B.T.M.K.), the ARIEL project of the Pediatric AIDS Foundation (B.T.M.K. and S.M.W.), and Public Health Service grant 5-U01-A135039 from the National Institutes of Health (S.M.W.).

REFERENCES

1. **Andeweg, A. C., P. Leeflang, A. D. M. E. Osterhaus, and M. L. Bosch.** 1993. Both the V2 and V3 regions of the human immunodeficiency virus type 1 surface glycoprotein functionally interact with other envelope regions in syncytium formation. J. Virol. **67:** 3232–3239.

2. **Berger, E. A., J. R. Sisler, and P. L. Earl.** 1992. Human immunodeficiency virus type 1 envelope glycoprotein molecules containing membrane fusion-impairing mutations in the V3 region efficiently undergo soluable CD4-stimulated gp120 release. J. Virol. **66:** 6208–6212.
3. **Bergeron, L., N. Sullivan, and J. Sodroski.** 1992. Target cell-specific determinants of membrane fusion within the human immunodeficiency virus type 1 gp120 third variable region and gp41 amino terminus. J. Virol. **66:**2389–2397.
4. **Blahut, R. E.** 1987. Information theory and statistics. Addison-Wesley, Reading, Mass.
5. **Boyd, M. T., G. R. Simpson, A. J. Cann, M. A. Johnson, and R. A. Weiss.** 1993. A single amino acid substitution in the V1 loop of human immunodeficiency virus type 1 gp120 alters cellular tropism. J. Virol. **67:**3649–3652.
6. **Brew, B. J.** 1993. HIV-1-related neurological disease. J. Acquired Immune Defic. Syndr. **6**(Suppl. 1):S10–S15.
7. **Budka, H., C. A. Wiley, P. Kleihues, J. Artigas, A. K. Asbury, E. S. Cho, D. R. Cornblath, M. C. D. Canto, U. DeGirolami, D. Dickson, L. G. Epstein, M. M. Esiri, F. Giangaspero, G. Gosztonyi, F. Gray, J. W. Griffin, D. Hennin, Y. Iwasaki, R. S. Janssen, R. T. Johnson, P. L. Lantos, W. D. Lyman, J. C. McArthur, K. Nagashima, N. Peress, C. K. Petito, R. W. Price, R. H. Rhodes, M. Rosenblum, G. Said, F. Scaravilli, L. R. Sharer, and H. V. Vinters.** 1991. HIV-associated disease of the nervous system: review of nomenclature and proposal for neuropathology-based terminology. Brain Pathol. **1:**143–152.
8. **Cheng-Mayer, C., J. Homsy, L. A. Evans, and J. A. Levy.** 1988. Identification of human immunodeficiency virus subtypes with distinct patterns of sensitivity to serum neutralization. Proc. Natl. Acad. Sci. USA **85:**2815–2819.
9. **Cheng-Mayer, C., J. T. Rutka, M. L. Rosenblum, T. McHugh, D. P. Sites, and J. A. Levy.** 1987. Human immunodeficiency virus can productively infect cultured human glial cells. Proc. Natl. Acad. Sci. USA **84:**3526–3530.
10. **Cheng-Mayer, C., C. Weiss, D. Seto, and J. A. Levy.** 1989. Isolates of human immunodeficiency virus type 1 from the brain may constitute a special group of the AIDS virus. Proc. Natl. Acad. Sci. USA **86:**8575–8579.
11. **Chesebro, B., K. Wehrly, J. Nishio, and S. Perryman.** 1992. Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell tropic isolates: definition of critical amino acids involved in cell tropism. J. Virol. **66:**6547–6554.
12. **Coffin, J. M.** 1992. Genetic diversity and evolution of retroviruses, Curr. Top. Microbiol. Immunol. **176:**143–164.
13. **de Jong, J. J., A. de Ronde, W. Keulen, M. Tersmette, and J. Goudsmit.** 1992. Minimal requirements for the human immunodeficiency virus type 1 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. J. Virol. **66:**6777–6780.
14. **Delassus, S., R. Cheynier, and S. Wain-Hobson.** 1992. Nonhomogeneous distribution of human immunodeficiency virus type 1 proviruses in the spleen. J. Virol. **66:**5642–5645.
15. **Dirks, R. W., R. P. M. van Gijlswilk, R. H. Tullis, A. B. Smit, J. Van Minnen, M. Van der Ploeg, and A. K. Rapp.** 1990. Simultaneous detection of different mRNA sequences coding for neuropeptide hormones by double *in situ* hybridization using FITC- and biotin-labeled oligonucleotides. J. Histochem. Cytochem. **38:** 467–473.
16. **Dirks, R. W., R. P. M. van Gijlswilk, M. A. Vooljs, A. B. Smit, J. Bogerd, J. Van Minnen, A. K. Rapp, and M. Van der Ploeg.** 1991. 3'-End fluorochromized and haptenized oligonucleotide as *in situ* hybridization probes for multiple, simultaneous RNA detection. Exp. Cell Res. **194:**310–315.
16a. **Douglas, S.** Unpublished data.
16b. **Dubois-Dolcq, M.** Unpublished data.
17. **Embretson, J., M. Zupancic, J. Beneke, M. Till, S. Wolinsky, J. Ribas, A. Burke, and A. Haase.** 1993. Analysis of human immunodeficiency virus-infected tissues by amplification and in situ hybridization reveals latent and permissive infections at single cell resolution. Proc. Natl. Acad. Sci. USA **90:**357–361.
18. **Epstein, L. G., C. Kuiken, B. M. Blumberg, S. Hartman, L. R. Sharer, M. Clement, and J. Goudsmit.** 1991. HIV-1 V3 domain variation in brain and spleen of children with AIDS: tissue-specific evolution within host determined quasispecies. Virology **180:**583–590.
19. **Faulkner, D. V., and A. Jurka.** 1988. Multiple aligned sequence editor (MASE). Trends Biochem. Sci. **13:**321–322.
20. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:**783–791.
21. **Felsenstein, J.** 1989. PHYLIP—phylogeny inference package (version 3.2). Cladistics **5:**164–166.
22. **Felsenstein, J.** 1993. PHYLIP (phylogeny inference package) version 3.4. Department of Genetics, University of Washington, Seattle. Distributed by the author.
23. **Fouchier, R. A. M., M. Groenink, N. A. Kootstra, M. Tersmette, H. G. Huisman, F. Miedema, and H. Schuitemaker.** 1992. Phenotype-associated sequence variation in the third variable region of the human immunodeficiency virus type 1 gp120 molecule. J. Virol. **66:**3183–3187.
24. **Freed, E. O., D. J. Myers, and R. Risser.** 1991. Identification of the principal neutralizing determinant of human immunodeficiency virus type 1 as a fusion domain. J. Virol. **65:**190–194.
25. **Freed, E. O., and R. Risser.** 1991. Identification of conserved residues in the human immunodeficiency virus types 1 principal neutralizing determinant that are involved in fusion. AIDS Res. Human Retroviruses **7:**807–811.
26. **Gartner, S., P. Markovits, D. M. Markovitz, M. H. Kaplan, R. C. Gallo, and M. Popovic.** 1986. The role of mononuclear phagocytes in HTLV-III/LAV infection. Science **233:**215–219.
27. **Goldstein, G. W., and A. L. Betz.** 1986. The blood-brain barrier. Sci. Am. **255:**74.
28. **Groenink, M., R. A. M. Fouchier, S. Broersen, C. H. Baker, M. Koot, A. B. van't Wout, H. G. Huisman, F. Miedema, M. Tersmette, and H. Schuitemaker.** 1993. Relation of phenotype evolution of HIV-1 to envelope V2 configuration. Science **260:**1513–1516.
29. **Hillis, D. M., M. W. Allard, and M. M. Miyamoto.** 1993. Analysis of DNA sequence data: phylogenetic inference. Methods Enzymol. **224:**456–487.
30. **Hillis, D. M., and J. J. Bull.** 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. **42:**182–189.
31. **Holmes, E. C., L. Q. Zang, P. Simmonds, C. A. Ludlam, and A. J. Leigh Brown.** 1992. Convergent and divergent sequence evolution in the surface envelope glycoprotein of the human immunodeficiency virus type 1 within a single patient. Proc. Natl. Acad. Sci. USA **89:**4835–4839.
32. **Hwang, A. S., and J. M. Coffin.** 1992. How does variation count? Nature (London) **359:**107–108.
33. **Hwang, S. S., T. J. Boyle, H. K. Lyerly, and B. R. Cullen.** 1991. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. Science **253:**71–74.
34. **Jordan, C., B. Watkins, C. Kufta, and M. Dubois-Dacq.** 1991. Infection of brain microglial cells by human immunodeficiency virus type 1 is CD4 dependent. J. Virol. **65:**736–742.
35. **Keys, B., J. Karis, B. Fadeel, A. Valentine, G. Norkrans, L. Hagberg, and F. Chioda.** 1993. V3 sequences of paired HIV-1 isolates from blood and cerebrospinal fluid cluster according to host and show variation related to the clinical stage of disease. Virology **196:**475–483.
36. **Koenig, S., H. Gendelman, J. Orenstein, M. D. Canto, G. Pezeshkpour, M. Yungbluth, F. Janotta, A. Aksamit, M. Martin, and A. Fauci.** 1986. Detection of AIDS virus in macrophages in brain tissue from AIDS patients with encephalopathy. Science **233:** 1089–1093.
37. **Korber, B., and G. Myers.** 1992. Signature patterns analysis: a method for assessing viral sequence relatedness. AIDS Res. Human Retroviruses **8:**1549–1558.
38. **Korber, B., S. Wolinsky, B. Haynes, K. Kuntsman, R. Levy, M. Furtado, P. Otto, and G. Myers.** 1992. HIV-1 intrapatient sequence diversity in the immunogenic V3 region. AIDS Res. Human Retroviruses **8:**1461–1465.
39. **Korber, B. T. M., R. M. Farber, D. H. Wolpert, and A. S. Lapedes.** 1993. Covariation of mutations in the V3 loop of human immu-

nodeficiency virus type 1 envelope protein: an information theoretic analysis. Proc. Natl. Acad. Sci. USA **90**:2176–2180.

40. **Kowalski, M., L. Bergeron, T. Dorfman, W. Haseltine, and J. Sodroski.** 1991. Attenuation of human immunodeficiency virus type 1 cytopathic effect by a mutation affecting the transmembrane envelope glycoprotein. J. Virol. **65**:281–291.
41. **Koyanagi, Y., S. Miles, R. T. Mitsuyasu, J. E. Merrill, H. V. Vinters, and I. S. Y. Chen.** 1987. Dual infection of the central nervous system by AIDS viruses with distinct cellular tropisms. Science **236**:819–822.
42. **Kuiken, C. L., J. J. de Jong, E. Baan, W. Keulen, M. Tersmette, and J. Goudsmit.** 1992. Evolution of the V3 envelope domain in proviral sequences and isolates of human immunodeficiency virus type 1 during transition of virus biological phenotype. J. Virol. **66**:4622–4627.
43. **Kuiken, C. L., and B. T. M. Korber.** 1994. Epidemiological significance of intra- and inter-person variation of HIV-1. AIDS **8**:S73–S83.
44. **Kuiken, C. L., G. Zwart, E. Baan, R. A. Coutinho, J. A. R. van den Hoek, and J. Goudsmit.** 1993. Increasing antigenic and genetic diversity of the HIV-1 V3 domain in the course of the AIDS epidemic. Proc. Natl. Acad. Sci. USA **90**:9061–9065.
45. **Kusumi, K., B. Conway, S. Cunningham, A. Berson, C. Evans, A. K. N. Iversen, D. Colvin, M. V. Gallo, S. Coutre, E. G. Shpaer, D. V. Faulkner, A. deRonde, S. Volkman, C. Williams, M. S. Hirsch, and J. I. Mullins.** 1992. Human immunodeficiency virus type 1 envelope gene structure and diversity in vivo and after cocultivation in vitro. J. Virol. **66**:875–885.
46. **Lee, W.-R., W.-J. Syu, B. Du, M. Matsuda, S. Tan, A. Wolf, M. Essex, and T.-H. Lee.** 1992. Nonrandom distribution of gp120 N-linked glycosylation sites important for infectivity of human immunodeficiency virus type 1. Proc. Natl. Acad. Sci. USA **89**:2213–2217.
47. **Leonard, C. K., M. W. Spellman, L. Riddle, R. J. Harris, J. N. Thomas, and T. J. Gregory.** 1990. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. J. Biol. Chem. **265**:10373–10381.
48. **Levy, R. M., E. R. Russell, M. Yungbluth, D. F. Hidvegi, B. A. Brody, and M. C. Dal Canto.** 1992. The efficacy of image-guided stereotactic brain biopsy in neurologically symptomatic acquired immunodeficiency syndrome patients. Neurosurgery **30**:186–190.
49. **Li, X. L., T. Moudgil, H. V. Vinters, and D. D. Ho.** 1990. CD4-independent, productive infection of a neuronal cell line by human immunodeficiency virus type 1. J. Virol. **64**:1363–1387.
50. **Li, Y., J. C. Kappes, J. A. Conway, R. W. Price, G. M. Shaw, and B. H. Hahn.** 1991. Molecular characterization of human immunodeficiency virus type 1 cloned directly from uncultured human brain tissue: identification of replication-competent and -defective viral genomes. J. Virol. **65**:3973–3985.
51. **Li, Y., L. Luo, N. Rasool, and C. Y. Kang.** 1993. Glycosylation is necessary for the correct folding of human immunodeficiency virus gp120 in CD4 binding. J. Virol. **67**:584–588.
52. **McKeating, J. A., J. Cordell, C. J. Dean, and P. Balfe.** 1992. Synergistic interactions between ligands binding to the CD4 binding site and V3 domain of human immunodeficiency virus type 1 gp120. Virology **191**:732–742.
53. **McNearney, T., P. Westervelt, B. J. Thielan, D. B. Trowbridge, J. Garcia, R. Whittier, and L. Ratner.** 1990. Limited sequence heterogeneity among biologically distinct human immunodeficiency virus type 1 isolates from individuals involved in a clustered infectious outbreak. Proc. Natl. Acad. Sci. USA **87**:1917–1921.
54. **Milich, L., B. Margolin, and R. Swanstrom.** 1993. V3 loop of the human immunodeficiency virus type 1 Env protein: interpreting sequence variability. J. Virol. **67**:5623–5634.
55. **Moore, J. P., and P. L. Nara.** 1991. The role of the V3 loop of gp120 in HIV infection and binding. AIDS **5**:s31–s33.
56. **Myers, G., B. Korber, S. Wain-Hobson, R. F. Smith, and G. N. Pavlakis.** 1993. Human retroviruses and AIDS 1993. Los Alamos National Laboratory for Theoretical Biology and Biophysics, Los Alamos, N. Mex.
57. **Myers, G., and R. Lenroot.** 1992. HIV glycosylation: what does it portend? AIDS Res. Human Retroviruses **8**:1459–1460.
58. **O'Brian, W. A., Y. Koyanagi, A. Namazie, J. Zhao, A. Diagne, K. Idler, and I. S. Y. Chen.** 1990. HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120 which do not include the CD4 binding domain. Nature (London) **348**:69–73.
59. **Pang, S., H. V. Vinters, T. Akashi, W. A. O'Brien, and I. S. Y. Chen.** 1991. HIV-1 env sequence variation in brain tissue of patients with AIDS-related neurologic disease. J. Acquired Immune Defic. Syndr. **4**:1082–1092.
60. **Pinter, A., W. J. Honnen, and S. A. Tilley.** 1993. Conformational changes affecting the V3 and CD4-binding domains of human immunodeficiency virus type 1 gp120 associated with *env* processing and with binding of ligands to these sites. J. Virol. **67**:5692–5697.
61. **Sattentau, Q. J., and J. P. Moore.** 1991. Conformational changes in the human immunodeficiency virus envelope glycoproteins by soluable CD4 binding. J. Exp. Med. **174**:407–415.
62. **Sharer, L. R.** 1992. Pathology of HIV-1 infection of the central nervous system. J. Neuropathol. Exp. Neurol. **51**:3–11.
63. **Sharpless, N. E., W. A. O'Brien, E. Verdin, C. V. Kufta, I. S. Chen, and M. Dubois-Dalcq.** 1992. Human immunodeficiency virus type 1 tropism for brain microglial cells is determined by a region of the *env* glycoprotein that also controls macrophage tropism. J. Virol. **66**:2588–2593.
64. **Shioda, T., J. A. Levy, and C. Cheng-Mayer.** 1992. Small amino acid changes in the V3 hypervariable region of gp120 can affect the T-cell-line and macrophage tropism of human immunodeficiency virus type 1. Proc. Natl. Acad. Sci. USA **89**:9434–9438.
65. **Smith, R. F., and T. F. Smith.** 1990. Automatic generation of primary sequence patterns from sets of related protein sequences. Proc. Natl. Acad. Sci. USA **87**:118–122.
66. **Smith, R. F., and T. F. Smith.** 1992. Pattern-induced multi-sequence alignment (PIMA) algorithm employing secondary structure-dependent gap penalties for use in comparative protein modeling. Protein Eng. **5**:35–41.
67. **Stamatatos, L., and C. Cheng-Mayer.** 1993. Evidence that the structural conformation of envelope gp120 affects human immunodeficiency virus type 1 infectivity, host range, and syncytium-forming ability. J. Virol. **67**:5635–5639.
68. **Steuler, H., B. Storch-Hagenlocher, and B. Wildemann.** 1992. Distinct populations of human immunodeficiency virus type 1 in blood and cerebrospinal fluid. AIDS Res. Human Retroviruses **8**:53–59.
69. **Swofford, D.** 1993. PAUP (phylogenetic analysis using parsimony) version 3.1. Center for Biodiversity, Illinois Natural History Survey, Champaign, Ill.
70. **Swofford, D. L., and G. J. Olson.** 1991. Phylogeny reconstruction, p. 411–566. *In* D. M. Hillis and C. Moritz (ed.), Molecular systematics. Sinauer Associates, Inc., Sunderland, Mass.
71. **Tornatore, C., A. Nath, K. Amemiya, and E. O. Major.** 1991. Persistent human immunodeficiency virus type 1 infection in human fetal glial cells reactivated by T-cell factor(s) or by the cytokines tumor necrosis factor alpha and interleukin-1 beta. J. Virol. **65**:6094–6100.
72. **Vartanian, J. P., A. Meyerhans, B. Asjo, and S. Wain-Hobson.** 1991. Selection, recombination, and G→A hypermutation of human immunodeficiency virus type 1 genomes. J. Virol. **65**:1779–1788.
73. **Wain-Hobson, S.** 1989. HIV genome variability in vivo. AIDS **3**(Suppl. 1):S13–S18.
74. **Watkins, B. A., H. H. Dorn, W. B. Kelly, R. C. Armstrong, B. Potts, F. Michaels, C. V. Kufta, and M. Dubois-Dulcq.** 1990. Specific tropism of HIV-1 for microglial cells in primary human brain cultures. Science **249**:549–553.
75. **Weber, J., P. Clapham, J. McKeating, M. Stratton, E. Robey, and R. Weiss.** 1989. Infection of brain cells by diverse human immunodeficiency virus isolates: role of CD4 as receptor. J. Gen. Virol. **70**:2653–2260.
76. **Westervelt, P., H. E. Gendelman, and L. Ratner.** 1991. Identification of a determinant within the human immunodeficiency virus type 1 surface envelope glycoprotein critical for productive infection of primary monocytes. Proc. Natl. Acad. Sci. USA **88**:3097–3101.

77. **Westervelt, P., D. B. Trowbridge, L. G. Epstein, B. M. Blumberg, Y. Li, B. H. Hahn, G. M. Shaw, R. W. Price, and L. Ratner.** 1992. Macrophage tropism determinants of human immunodeficiency virus type 1 in vivo. J. Virol. **66:**2577–2582.
78. **Wiley, C., R. Schrier, J. Nelson, P. Lampert, and M. Oldstone.** 1986. Cellular localization of HIV infection within brains of AIDS patients. Proc. Natl. Acad. Sci. USA **83:**7089–7093.
79. **Wolfs, T. F. W., G. Zwart, M. Bakker, and J. Goudsmit.** 1992. HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. Virology **189:**103–110.
80. **Wolinsky, S. M., C. M. Wike, B. Korber, C. Hutto, W. P. Parks, L. L. Rosenblum, K. J. Kunstman, M. R. Furtado, and J. Munoz.** 1992. Selective transmission of human immunodeficiency virus type 1 variants from mothers to infants. Science **255:**1134–1137.
81. **Wyatt, R., M. Thali, S. Tilley, A. Pinter, M. Posner, D. Ho, J. Robinson, and J. Sodroski.** 1992. Relationship of the human immunodeficiency virus type 1 gp120 third variable loop to a component of the CD4 binding site in the fourth conserved region. J. Virol. **66:**6997–7004.
82. **Zhang, L. Q., P. MacKenzie, A. Cleland, E. C. Holmes, A. J. Leigh Brown, and P. Simmonds.** 1993. Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. J. Virol. **67:**3345–3356.
83. **Zhu, T., H. Mo, N. Wang, D. S. Nam, Y. Cao, R. A. Koup, and D. D. Ho.** 1993. Genotypic and phenotypic characterization of HIV-1 in patients with primary infection. Science **261:**1179–1181.